Using Paradata to Understand Effort and Attrition in a Panel Survey

Andrew Mercer[1]
[1]Westat, 1600 Research Boulevard, Rockville, MD 20850

Abstract

Current practices in survey research prioritize high response rates as the key indicator of survey quality. Survey organizations often expend large amounts of time and resources to attain responses from more difficult to interview respondents, particularly those that require refusal conversion, tracking or are otherwise hard to reach. Longitudinal surveys face the additional challenge of maintaining a high response rate over multiple rounds in order to minimize the effects of attrition. In this paper, we use contact record paradata from the 2009 cohort of the Medical Expenditure Panel Survey in order to construct three scenarios that simulate the effects of restricting recruitment effort spent on locating, contacting and obtaining cooperation from respondents respectively. For each of these scenarios, we examine differences in the pattern of attrition over the course of the survey and the changes that occur in estimates of key indicators as a result. For this survey, we find that restricting contacting effort results in the largest increase in overall attrition. Reducing effort put toward locating and contacting are found to produce sizeable changes in the resulting estimates. Reducing refusal conversion effort produces a comparable level of additional nonresponse to locating effort, however the impact on estimates is minimal.

Key Words: Paradata, longitudinal survey, nonresponse, attrition, survey effort

1. Introduction

Standard survey best practices emphasize maximizing response rates in order to minimize the risk of nonresponse bias. To this end, survey organizations devote tremendous resources toward obtaining as many complete interviews as possible. Such efforts include the use of pre-notification letters, monetary incentives, repeated phone calls and in-person visits, and the employment of traveling refusal conversion specialists. Despite these efforts, there is limited evidence of a strong correlation between nonresponse rates and nonresponse bias (Groves and Peytcheva 2008). It is reasonable to investigate the feasibility of reducing the costs associated with achieving maximal response rates provided the tradeoffs with respect to increased measurement error or nonresponse bias are within acceptable ranges. In recent years, a body of methodological research has explored the extent to which variations in the level of recruitment effort affect survey estimates.

To date, these analyses have primarily focused on cross-sectional surveys, while panel surveys remain largely unexamined. Although panel surveys and cross-sectional surveys share many of the same features, panel surveys require some additional considerations with respect to effort expended. While the first round of a panel survey is identical to a cross-sectional survey in most respects, the necessity of maintaining respondent

cooperation over the course of multiple rounds may have implications beyond the possibility of bias in the initial round. In any panel survey, some amount of attrition is bound to occur. Design decisions that limit the number of completed cases in Round 1 determine an initial sample size that will only diminish over subsequent rounds. Restrictions on effort expended in subsequent rounds may compound, reducing the final sample size and possibly magnifying attrition bias. On the other hand, it is possible that respondents with a high propensity for attrition simply drop out earlier in the process with minimal effect on the final sample composition. This analysis seeks to expand upon the existing body of research on level of effort by examining how variations in effort affect the both rate of attrition and resulting survey estimates.

The level of nonresponse in a survey does not automatically result in bias. Rather, the degree of bias depends on the degree to which nonrespondents differ from respondents for any given statistic (Groves 2004). As such, nonresponse bias is not uniform across survey items, but will vary for each estimate. Although maintaining high response rates is generally seen as a guard against the risk of nonresponse bias, Groves and Peytcheva (2008) demonstrated that substantial nonresponse bias is possible even when response rates are high. Restricting the number of allowed contact attempts or limiting refusal conversion efforts will result in higher nonresponse among respondents who are difficult to contact or uncooperative. If these hard to reach respondents are systematically different from easier to reach respondents bias would result.

In 1997, Scott Keeter and colleagues fielded two identical RDD telephone surveys (Keeter et al. 2000). One applied the Pew Research Center's standard methodology, which limited the field period to five days, employed a less rigorous screening method, and allowed only one refusal conversion attempt. They compared this to a more rigorous study that fielded over the course of eight weeks and employed more thorough attempts to locate and obtain cooperation from respondents including an advance letter and $2 incentive. The standard survey obtained a response rate of 36% compared to 61% for the more rigorous survey. Despite the sizeable difference in response rate, they found few significant differences between survey estimates. This experiment was replicated again in 2003 and again found few significant differences in estimates (Keeter et al. 2006). Similarly, Curtin et al. (2000) generated estimates of the Index of Consumer Sentiment using data from 211 monthly iterations of the Survey of Consumer Attitudes after removing from the sample cases that required two or more and five or more calls to complete. They found evidence of systematic differences in the estimates when the effort was artificially limited; however the effects were generally small and only statistically significant when many surveys were pooled into unusually large samples. These studies demonstrate that even severe reductions in response rate due to restricted effort do not necessarily result in equally severe changes to estimates.

Where differences have been found, much of the research related to recruitment effort has focused on the quality of the responses given by difficult respondents. Cannell and Fowler (1963) first hypothesized that respondents lacking in motivation would require disproportionate effort to interview and provide lower quality responses. In a review of the literature, Olson (forthcoming) found evidence that respondents who are difficult to contact or reluctant to participate yield higher levels of item nonresponse, particularly for items that are sensitive or difficult, though evidence that these respondents provide less accurate or reliable answers is limited. Survey organizations facing budgetary constraints would be thrilled to confirm that reducing interviewer effort simultaneously reduces costs and measurement error without significantly increasing nonresponse bias. Empirical

results are not so clear-cut. Kreuter et al. (2010) were able to link both contact record paradata and administrative records to survey estimates of welfare benefits, employment, age and citizenship in a comprehensive assessment of the nonresponse bias and measurement error present at different levels of effort. They found that increased measurement error among difficult respondents was minimal, while nonresponse bias was significantly reduced with the inclusion of hard to interview respondents.

In cross-sectional surveys, obtaining a completed interview requires locating the respondent, successfully making contact, and obtaining cooperation. Failure at any one of these tasks will result in an inability to complete an interview. In panel surveys, the process is the same; however it must be repeated at each round. Lepkowski and Couper (2002) theorize that a respondent's overall response propensity for any given round is a function of their propensities for location, contact and cooperation conditional upon respondent characteristics, survey design features, organizational efforts and experiences during prior rounds of interviewing. Each of these tasks requires different kinds of effort on the part of the interviewer and survey organization. Locating respondents who have moved can involve conducting internet or public record searches, contacting former neighbors and relatives, or knocking on doors of possible new residencies. Contacting effort can include phone calls, in person visits. Further, if a respondent has relocated, such effort is wasted if it is not immediately apparent that the respondent has moved. Refusal conversion effort can involve recontacting respondents, providing monetary incentives, and sometimes sending interviewers with a specialty in refusal conversion from outside of the area. Reducing effort expended on any of these three components would reduce response propensity for individuals and increase attrition in the aggregate.

There is no reason to assume that these three mechanisms affect survey estimates in the same way. In a cross-sectional context, Lynn et al. (2002) found that respondents who are difficult to contact showed greater differences from the remaining cases than did respondents who were uncooperative, although significant differences were apparent for both groups. Lepkowski and Couper (2002) found that in the Americans' Changing Lives panel survey, certain respondent characteristics such as employment status, financial difficulty and residential mobility were associated with ease of location but not cooperation in the second round. Prior round survey experience was strongly correlated with cooperation but not location, while measures of community attachment affected both location and cooperation. Watson and Wooden (2009) found that correlates of attrition due to failure to contact are distinct from those related to cooperation. For example, renting a home, marital status and the presence of children were all associated with contact propensity but not cooperation. In their study, employment was positively associated with contact but negatively associated with cooperation. Couper and Ofstedal (2009) found that tracking and locating was more likely to be required for racial minorities, lower educated and younger respondents. Contrary to expectations, they found that renters and low-income households were not significantly associated with requiring location when other covariates are controlled for. Other analyses of attrition have resulted in similar findings (see Nicoletti and Peracchi 2005, Lemay 2009). There is substantial evidence that propensities for location, contact and cooperation vary between individuals and across survey rounds.

This analysis is an initial attempt to relate interviewer effort toward locating, contacting, and obtaining cooperation on survey estimates and sample composition over the course of a panel survey. To this end, two research questions present themselves as good starting points for analysis:

- How do restrictions on recruitment effort affect the pattern of attrition? Do restrictions in effort compound nonresponse over the course of a panel survey, or do respondents with a higher attrition propensity simply drop out earlier in the process? Does this vary according to the type of effort that is restricted?
- What are the effects of separately restricting effort toward locating, contacting, and obtaining cooperation on sample composition and key substantive indicators both in individual rounds and over the course of a survey?

To estimate these effects, we employ a simulation strategy similar to that of Curtin et al. (2000). Using both paradata and interview data from the 2009 panel of the Medical Expenditure Panel Survey (MEPS), we will test the relative effects of post hoc restrictions on level of effort toward locating, contacting, and gaining cooperation from respondents. Respondents who exceed a specified effort threshold in a given round are removed from the sample for that round and all subsequent waves. Although not a true counterfactual, this approach allows us to calculate estimates similar to those that would have resulted under the specified conditions, and compare these estimates to the original estimates in the absence of these restrictions.

2. Data

MEPS is an ongoing, nationally representative face-to-face survey of households in the United States. The survey tracks Americans' health conditions, use of medical services and medical expenditures, as well as economic and demographic indicators. MEPS is conducted by the Agency for Healthcare Research and Quality and the National Center for Health Statistics. In addition to a household component, the survey also features a medical provider component and a pharmacy component where medical and prescription records are obtained for individuals who consent. In this analysis, only the household component is employed.

MEPS uses an overlapping panel design. Each year, a new panel is selected from a frame consisting of households that completed the prior year's National Health Interview Survey (NHIS). The NHIS uses a stratified, multi-stage area probability sample design that oversamples Hispanic, Black, and Asian households. This complex design carries through to MEPS. This analysis uses MEPS Panel 14 which included 9,899 households that completed the 2008 NHIS. Respondents are considered in scope for two years; the reference period for Panel 14 began in January 2009 and continued through December 2010. Respondents are interviewed a total of five times approximately six months apart. In Panel 14, respondents received an advance mailing notifying them of their selection and including a $5 incentive. Respondents also received another mailing prior to each round, and an additional $30 upon completion of the interview for each round (Westat 2011).

The MEPS sample is drawn at the dwelling unit level. Dwelling units can contain one or more reporting unit (RU), which is roughly analogous to a family. In MEPS, most dwellings contain only one RU, although dwellings with unrelated roommates or multiple families contain more than one. Some RUs split up in situations such as divorce, or a child moving out, at which point a new RU is created. Under these circumstances, certain RUs do not come into existence until later rounds of data collection. We restrict the present analysis to only the 8,559 RUs that did not share a dwelling with other RUs, and that did not split up during the time they were under observation. This is done in order to ensure that the units of analysis remain consistent and to prevent analytic problems

associated with left-censored observations, a situation that occurs when the beginning of observation begins later for some units than others. One additional RU was removed from consideration because paradata indicated that work continued on that case after it had already been identified as having dropped out of the survey.

Estimates of the level of effort for any given case are derived from activity log paradata contained in the survey's case management system. Interviewers are expected to create an entry in this database for each action they take in working a case. Each record indicates the date the action was taken as well as the nature of the action, mode of contact and a disposition indicating the nature and outcome of the action. These records also indicate whether or not the action resulted in successful contact with a member of the household who is eligible to complete the interview. By looking at activity records for each case, we can determine how many actions were required to complete an interview, whether or not a household moved and required location and tracking, and whether a case required refusal conversion. These activity log records allow us to identify if and when households drop out of panel and the level and types of effort that were expended while they were under observation.

The availability of activity log paradata enables us to construct several alternative sets of estimates in order to determine the effects of restricting effort related to locating, contact and refusal conversion respectively. The baseline scenario consists of the 8,558 households identified above. These households are used to construct estimates of the hazard function for attrition at each round of data collection. Data for the 12,460 individual respondents who belong to these households and remained in scope for all five rounds are used to calculate estimates for key demographic and health related indicators such as prevalence of cancer and insurance coverage. Restricting estimates to only those respondents who were in scope for the full duration of the survey eliminates from consideration individuals who were institutionalized or began serving in the military, as well as individuals who were either born or died during the field period. These individuals share important characteristics with respect to health conditions and medical expenditures, particularly the deceased and newborn infants (Machlin and Yu 2005). However, removing these cases from consideration allows us to estimate the proportion of individuals who experienced a change in other aspects of their living situation such as their employment or marital status or insurance coverage. Using this dataset as a baseline for comparison, we then construct three alternative scenarios in which each type of effort, locating, contacting and refusal conversion respectively are assumed to have been restricted.

To devise a scenario consisting only of respondents who require no additional effort to locate, starting in Round 1 and proceeding through Round 5, any completed cases that were identified as requiring location and tracking efforts are removed from the sample in the round where the effort was required as well as from all subsequent rounds. For example, suppose a household completed Round 1 but moved prior to Round 2 and were unreachable through any of the contact information provided previously (such as on a cellular or work telephone number). Regardless of whether or not that case was successfully located and an interview was completed, we would treat that household as having dropped out of the survey during round 2, and remove any household members from the resulting individual dataset. This alternative dataset is an approximation of what would have resulted if location effort had never occurred. Such a maximal restriction on locating effort is somewhat unrealistic and would probably not be employed in a production survey. Rather it serves to indicate an upper bound on the level of attrition

that could result from restricting effort related to locating and tracking respondents, and provide a sense of the differences in estimates that would occur under such circumstances. This alternative scenario is referred to as the *locatable* scenario as it contains only those respondents who were consistently locatable throughout the data collection process.

To generate a dataset where the effort expended prior to first contact is restricted, referred to as the *contactable* scenario, we follow the same procedure as before, treating households that require more than four actions to make first contact with an eligible member of the household as dropping out of the sample during the round that this occurs. Although this threshold is somewhat arbitrary, this dataset enables the calculation of estimates under a stringent restriction on the level of effort expended to make initial contact with the household. It is important to note that actions are not the same as contact attempts. In a phone survey, call records generally only represent one kind of action, telephone calls. In a face-to-face survey, these records can contain numerous types of actions including telephone calls, personal visits, tracking activity such as database lookups, and mailing refusal conversion letters. Not all actions represent an equal expenditure of effort. In the absence of information on the number interviewer-hours spent per task, the number of actions serves as an imperfect but useful, catchall unit of effort that we assume balances out in the aggregate.

Table 1: Overview of Baseline and Alternative Scenarios

	Description	*Round 1 eligible RUs*	*Round 5 complete RUs*	*Person level obs.*
Full Survey	*MEPS Panel 14 with no exclusions or restrictions*	10,227	6,761	16,994
Baseline Scenario	*Excludes multi-RU dwellings and split RUs. Person-level estimates are restricted to individuals who were in-scope for 5 rounds.*	8,558	5,055	12,460
Locatable Scenario	*Same as baseline scenario but excludes RUs that required locating effort in any round.*	8,558	4,207	10,285
Contactable Scenario	*Same as baseline scenario but excludes RUs that required more than four actions to make contact in any round.*	8,558	3,455	8,546
Cooperative Scenario	*Same as baseline scenario but excludes RUs coded as having refused participation in any round.*	8,558	4,388	10,892

Finally, the *cooperative* scenario eliminates households that are coded as experiencing an interim-refusal prior to completion. Like the locatable scenario, this restriction precludes any efforts at refusal conversion. It must be noted that only interactions that ended with a refusal from the respondent would be coded as such. For instance, an interaction that began with the respondent refusing to be interviewed but ended with an interviewer convincing them to participate would not be coded as an interim refusal. Again, this serves to mark an upper bound on the level of attrition that can be attributed to restricted efforts to obtain cooperation as it does not draw a distinction between so-called “soft” and “hard” refusals. Table 1 summarizes the characteristics of each scenario to be considered.

3. Methods

Because MEPS employs a complex sample design with unequal probabilities of selection, standard statistical methods designed for simple random samples are inappropriate. Instead, point estimates must be calculated using weights that account for the probability of selection. When working with survey data, these weights typically include a nonresponse adjustment to correct for the over or under-representation of certain groups in the set of completed observations. In this case, we wish to compare the estimates that result under different nonresponse situations. The use of the standard nonresponse-adjusted weights would distort any comparison. Instead, we use base weights that only adjust for unequal probability of selection. Variance estimates are calculated using Fay's method of balanced repeated replication (BRR) (Judkins 1990). Because the reduced effort scenarios are simply subsets of respondents within the baseline scenario, estimating proportions and their standard errors is straightforward. However, statistical comparisons of these estimates are complicated because covariance between the scenarios cannot be determined analytically due to their partial overlap. Fortunately, replication allows us to get around this obstacle by calculating the difference of proportions separately for each replicate and averaging the squared deviation from the estimate for the full set of observations without the need to separately calculate the covariance.

To answer research question 1, we will compare the attrition pattern in the baseline scenario with the patterns found in the reduced effort scenarios. Attrition patterns are well suited for analysis using Kaplan-Meier survival and hazard functions (Singer and Willet 2003). Each round of the survey represents a discrete period of time t_j where $j = 1\ldots5$. Time $j = 0$ represents the sample prior to beginning of data collection before any attrition has occurred. The hazard function $h(t_j)$ represents probability that a household will drop out during the round designated as t_j conditional upon not having dropped out prior to the beginning of that round. In this context, the hazard rate for any given round can be represented as follows:

$$h\left(t_j\right) = \frac{d_j}{n_j}$$

Where d_j is the weighted number of households that drop out during round j, and n_j is the weighted number of households in the sample at the start of round j.

The survival function $S(t_j)$ represents the cumulative proportion of sampled households who remain in the survey up through and including time t_j. Because the event under consideration is failure to complete a round of data collection for any reason, there are no censored households other than those that complete all five rounds of data collection at which point observation ceases. In this analysis, $S(t_j)$ can be thought of as a simple form of cumulative response rate through round t_j. The survival function is represented mathematically as a function of the hazard rate:

$$S\left(t_j\right) = \prod_{1}^{j}\left[1 - h\left(t_j\right)\right]$$

To compare across scenarios, we construct an estimate of the difference between the hazard functions of the baseline and alternative scenarios. For example, the difference in hazards for the locatable scenario is represented by the equation:

$$\delta_{locatable,j} = h\left(t_j\right)_{locatable} - h\left(t_j\right)_{baseline}$$

This estimate δ measures the excess risk of attrition present in each round of data collection for each alternative scenario.

Although differences in attrition patterns among these scenarios tell us how many households may drop out of data collection and when it is likely to occur, they tell us nothing about the eventual impact on survey estimates. Effort protocols that result in higher nonresponse will not bias estimates if the additional nonresponse is distributed randomly throughout the population. In order to address research question 2, we will estimate the proportion of respondents belonging to key categories for the baseline and alternative scenarios and then estimate confidence intervals for the difference between the alternative scenarios and the baseline scenario. As an example, the estimated difference for men in the locatable scenario is calculated as:

$$\delta\left(male\right)_{locatable} = p\left(male\right)_{locatable} - p\left(male\right)_{baseline}$$

Estimates where zero falls within the confidence interval do not display a statistically significant difference between the baseline and reduced effort scenario.

4. Results

4.1 Effect on Attrition Patterns

For each scenario, the basic pattern of is consistent. Round 1 exhibits the highest risk of attrition, followed by a leveling off and gradual reduction in risk between rounds 2 and 5. Although the overall shape of the curves is similar, the survival-rate at round 5 is markedly lower in the reduced effort scenarios. The locatable and cooperative scenarios both end with 49% and 50% of households completing round 5 respectively, while in the contactable scenario only 40% of households complete round 5.

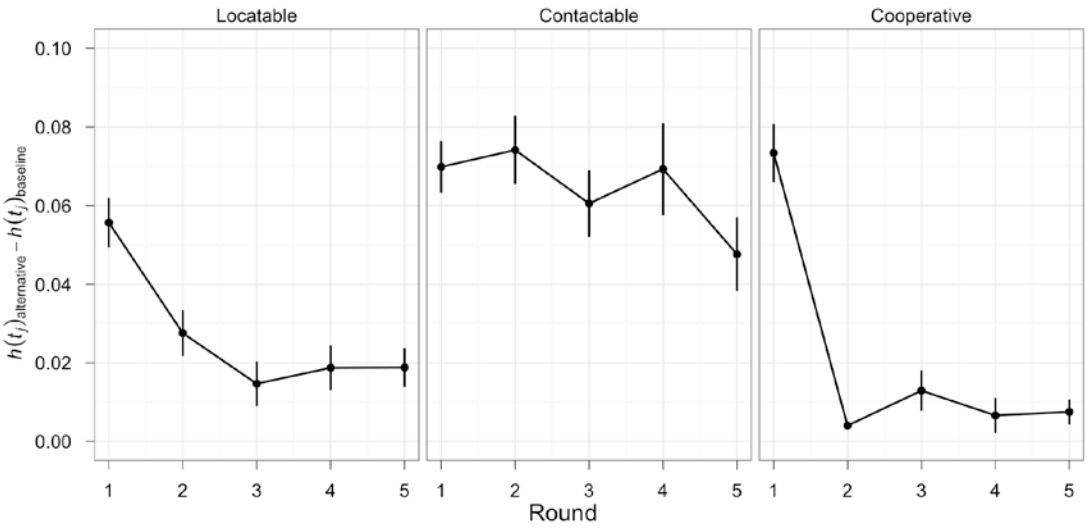

Figure 1: Comparison of Hazard Function in Reduced Effort Scenarios to Baseline Scenario

Figure 1 displays the difference between the alternative hazard rates and the baseline at each round of data collection. This represents the excess risk of attrition above and beyond the baseline for each round of data collection. Here we can see that in the locatable scenario, round 1 exhibits 5.5% greater risk of attrition than the baseline, which diminishes to 3% at round 2 before leveling off at 2%. This suggests that by round 3, transitory or mobile families have mostly worked their way out of the study. While the locatable scenario showed a reduction in added risk in rounds 1 and 2 before stabilizing, the contactable scenario shows only the slightest possible decrease in added risk with each subsequent round. Restricted contacting effort produces a consistent, steady drain on the completion rate at each round. Finally, in the cooperative scenario, we see that nearly all of additional the loss occurs in round 1. Among households that remain at round 2, the added risk diminishes to 1% and remains there for the duration of the survey.

Effect on Estimates

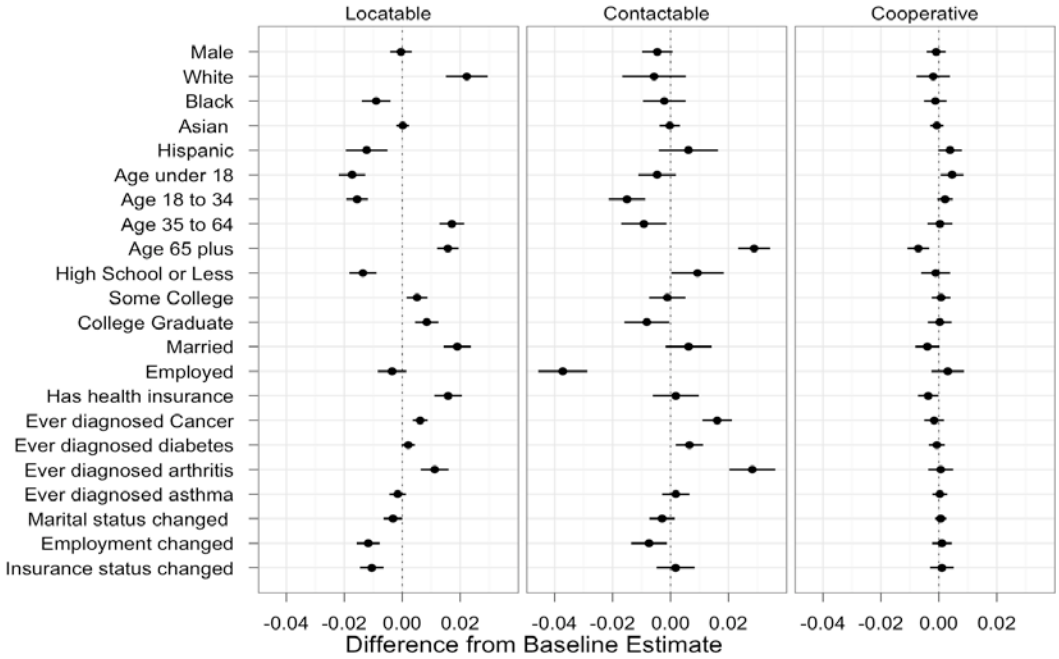

Figure 2: Difference Between Alternative Scenario and Baseline Estimates

Figure 2 displays the differences in proportions between the baseline scenario and the reduced effort scenarios, and their 95% confidence intervals. In the locatable scenario, where households that moved between rounds of data collection are excluded, whites are overestimated relative to the baseline while blacks and Hispanics are underestimated. Age is notably affected with younger respondents dropping out of the sample and older respondents remaining. Education displays a trend where respondents with a high school degree or less are underrepresented while higher education levels are overrepresented. The percent employed is largely unchanged. Among the substantive estimates, respondents with health insurance coverage are overrepresented by 0.02. Estimates of the prevalence of cancer and arthritis are also both positively affected by roughly 0.01. Locatable respondents exhibit fewer instances of change, particularly with respect to employment and health insurance coverage. Taken together these differences suggest that when locating effort is eliminated, the resulting sample is older, more educated, and stable. Higher prevalence of cancer and arthritis are likely explained by the increase in the proportion of older respondents. Households consisting of minorities, young people

and individuals with less education drop out of the survey. Households that relocate when a family member gains or loses a job are also more likely to be left out of these estimates.

When effort to make first contact is restricted, different patterns emerge. The proportion of respondents 65 years old or older is higher by 0.03 while the proportions for 18 to 34 and 35 to 64 years old are both significantly reduced. Here the effect on estimates of education levels is opposite from the locatable scenario, with the proportion having a high school degree or less overrepresented and college graduates underrepresented. The must substantial difference is the reduction in the proportion of employed by 0.04. These differences largely conform to what we would expect. Households with older respondents who are more likely to be retired and at home when an interviewer attempts to make contact are more likely to remain in the sample. Younger individuals and those who are employed are less likely to be at home and require more effort to make contact. Again, the higher proportion of elderly respondents occurs alongside higher estimates for cancer and arthritis. Prevalence of diabetes is also 0.01 higher than in the baseline scenario.

The cooperative scenario is perhaps most striking due to the absence of any clear patterns or notable deviations from the baseline scenario. There is a slight underrepresentation of respondents 65 or older relative to the baseline by about -0.01. The difference in the proportion who have health insurance is statistically significant, however the difference is only -0.004. None of the other differences reach statistical significance at a 95% confidence level. For the two estimates that are statistically different, the effect is quite small. Compared to locating and contact, eliminating households that required refusal conversion seems to produce a negligible effect on this set of survey estimates.

5. Discussion

The lack of any substantial effect from excluding respondents that required refusal conversion effort is particularly notable and suggests that much refusal conversion effort may be misdirected. However before research organizations begin slashing budgets for refusal conversion this phenomenon should be more thoroughly explored, preferably experimentally. It is possible that although few effects were found at the aggregate level, some important subgroups may be disproportionately affected. The differences related to contacting effort largely conform to expectations, although the finding that estimates of employment are lower when contacting effort is restricted contradicts Watson and Wooden's (2009) finding that contactability is positively associated with employment. It is possible that this difference is simply due to differences in operationalization or circumstances related to differences in the surveys under consideration. However such efforts in panel surveys represent a substantial investment of time and money. The findings associated with the locatable scenario are consistent with the earlier findings of Couper and Ofstedal (2009). The impact of reduced locating effort could be substantial, particularly with respect to correlates of age, education and minority status.

When considering these differences in estimates, two facts are important to bear in mind. First, although we observe differences, we cannot directly infer that the reduced effort scenarios produce bias because we do not know the true population parameter. It is possible that the baseline estimates, which are themselves subject to measurement error and nonresponse, are biased in the other direction. In such an instance, a reduced effort scenario could actually correct for bias already present in the baseline estimate. Nevertheless, we see that reductions in different kinds of effort can move estimates in

opposite directions, as is the case with education in the locatable and contactable scenarios. If key estimates are correlated with field protocols, the potential for bias exists.

Second, these changes in estimates are not simple functions of individual response propensities. In MEPS, nonresponse occurs at the household level, however these survey estimates are calculated at a person level. If an individual in a household gets a job in another state and relocates with the rest of his family, estimates in the locating scenario are changed not only by the loss of that individual, but also by the loss of every other household member.

Some of the differences that occur in these estimates such as age or education are easily correctable with a nonresponse weighting adjustment. Others such as the proportion who experienced a change in employment or medical insurance status may not be corrected by standard weighting adjustments. One avenue for future research would be to apply the standard MEPS nonresponse adjustment procedure to these reduced effort scenarios in order to assess the extent to which any differences can be corrected by weighting.

With respect to the patterns of attrition, this analysis has assumed that rules governing the level of effort expended do not differ by round. However, given the disproportionate level of nonresponse that occurs in round 1, it may be worthwhile to investigate different protocols for round 1 and subsequent rounds. For example, in the cooperative scenario, nearly all of the additional risk of nonresponse occurred in round one. It could be that allowing refusal conversion in round 1 but restricting it in subsequent rounds mitigates most of the nonresponse without substantially increasing attrition risk in subsequent rounds. Future analysis should incorporate a focus on the timing of effort in addition to the type and level.

Finally, the scenarios examined in this paper represent a greatly simplified version of a large-scale panel survey. This analysis excludes from consideration households with multiple RUs or RUs that split. However such households represent a substantial portion of the sample as a whole. Although they present a challenge for statistical analysis, a complete understanding of effort requires that they be taken into account. Despite these limitations, this serves as a valuable exercise in understanding the manner in which recruitment effort affects respondent retention over the course of a panel, and in turn how those dynamics have the potential to affect the estimates themselves. For a survey such as MEPS, where nonresponse occurs at the RU level and estimates are produced at the person level, this simulation approach allows us to gauge effects that would be difficult to estimate through modeling. It would be worthwhile to replicate this analysis with other surveys in order to ascertain the extent to which these findings are generalizable beyond the survey at hand.

References

Cannell, Charles F, and Floyd J Fowler. 1963. “Comparison of a Self-Enumerative Procedure and a Personal Interview: A Validity Study.” *Public Opinion Quarterly* 27(2):250–264.

Couper, Mick P, and Mary Beth Ofstedal. n.d. “Keeping in Contact with Mobile Sample Members.” Pp. 183–203 in *Methodology of Longitudinal Surveys*, edited by Peter Lynn. John Wiley & Sons, Ltd.

Curtin, Richard, Stanley Presser, and Eleanor Singer. 2000. "The Effects of Response Rate Changes on the Index of Consumer Sentiment." *Public Opinion Quarterly* 64(4):413–428.

Groves, Robert M, and Emilia Peytcheva. 2008. "The Impact of Nonresponse Rates on Nonresponse Bias A Meta-Analysis." *Public Opinion Quarterly* 72(2):167–189.

Groves, Robert M. 2004. *Survey Errors and Survey Costs*. Hoboken, NJ: Wiley-Interscience.

Judkins, David R. 1990. "Fay's Method for Variance Estimation." *Journal of Official Statistics* 6(3):223–239.

Keeter, S., C. Kennedy, M. Dimock, J. Best, and P. Craighill. 2006. "Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey." *Public Opinion Quarterly* 70(5):759–779.

Keeter, S., C. Miller, A. Kohut, R. M Groves, and S. Presser. 2000. "Consequences of Reducing Nonresponse in a National Telephone Survey." *Public Opinion Quarterly* 64(2):125–148.

Kreuter, F., G. Muller, and M. Trappmann. 2010. "Nonresponse and Measurement Error in Employment Research: Making Use of Administrative Data." *Public Opinion Quarterly* 74(5):880–906.

Lemay, Michael. 2009. "Understanding the Mechanism of Panel Attrition." Ph.D. dissertation, Joint Program in Survey Methodology, University of Maryland, College Park. Retrieved April 13, 2012 (http://drum.lib.umd.edu//handle/1903/9631).

Lepkowski, James M., and Mick P Couper. 2002. "Nonresponse in the Second Wave of Longitudinal Household Surveys." Pp. 259–272 in *Survey Nonresponse*, edited by Robert M Groves, Don A. Dillman, John L. Eltinge, and Roderick J.A. Little. Hoboken, NJ: John Wiley & Sons, Ltd.

Lynn, James M., Paul Clarke, Jean Martin, and Patrick Sturgis. 2002. "The Effects of Extended Interviewer Efrorts on Nonresponse Bias." Pp. 135–147 in *Survey Nonresponse*, edited by Robert M Groves, Don A. Dillman, John L. Eltinge, and Roderick J.A. Little. Hoboken, NJ: John Wiley & Sons, Ltd.

Machlin, Steven, and William Yu. 2005. "MEPS Sample Persons In-Scope for Part of the Year: Identification and Analytic Considerations." Agency for Healthcare Research and Quality, Rockville, MD. Retrieved April 13, 2012 (http://meps.ahrq.gov/mepsweb/about_meps/hc_sample.shtml).

Nicoletti, C., and F. Peracchi. 2004. "Survey response and survey characteristics: micro-level evidence from the European Commission Household Panel." CEIS Working Paper No. 64, Centre for Economic and International Studies, Rome, Italy.

Olson, Kristen. Forthcoming. "Do Nonresponse Follow-ups Improve or Reduce Data Quality? A Review of the Existing Literature." Journal of the Royal Statistical Society.

Singer, Judith D., and John B. Willett. 2003. Applied Longitudinal Data Analysis : Modeling Change and Event Occurrence. 1st ed. New York, NY: Oxford University Press, USA.

Watson, Nicole, and Mark Wooden. 2009. "Identifying Factors Affecting Longitudinal Survey Response." Pp. 157–181 in Methodology of Longitudinal Surveys, edited by Peter Lynn. Hoboken, NJ: John Wiley & Sons, Ltd.

Westat. 2011. "MEPS Annual Methodology Report." Agency for Healthcare Research and Quality, Rockville, MD. Retrieved April 13, 2012 (http://meps.ahrq.gov/mepsweb/data_files/publications/annual_contractor_report/hc_ann_cntrct_methrpt.shtml).